Математические и статистические методы в психологии
Проверка статистических гипотез. (6 декабря 2019 г.)
А. А. Макаров, А. А. Тамбовцева, Н. А. Василёнок, Е. П. Шеремет

Основные понятия

- **Нулевая гипотеза ($\boldsymbol{H_0}$)** – утверждение о параметре генеральной совокупности (параметрах генеральных совокупностей) или распределении, которое необходимо проверить.
- **Альтернативная гипотеза ($\boldsymbol{H_1}$)** – утверждение, противоположное нулевой гипотезе. Выдвигается, но не проверяется. На него мы «соглашаемся» в случае, если нулевую гипотезу на основе имеющихся данных необходимо отвернуть.

  Все гипотезы можно разделить на двусторонние (ненаправленные) и односторонние (направленные).

Двусторонние альтернативы	Односторонние альтернативы
$H_1 : \mu \neq 168$ (средний рост женщин не 168 см)	- левосторонние ($H_1 : \mu < 168$) - правосторонние ($H_1 : \mu > 168$)

- **Уровень значимости ($\boldsymbol{\alpha}$)** – вероятность отвергнуть верную нулевую гипотезу. Нулевая гипотеза всегда проверяется на определённом уровне значимости. Например, если мы проверяем нулевую гипотезу на уровне значимости 5%, это означает, что если мы будем проводить аналогичные исследования 100 раз и проверять на основе имеющихся данных интересующую нас нулевую гипотезу, в 5 случаях из 100 мы отвергнем нулевую гипотезу, хотя она будет верной.

  Уровень значимости в каком-то смысле является понятием, противоположным уровню доверия. Уровень доверия – вероятность не отвергнуть верную нулевую гипотезу. Проверить нулевую гипотезу на уровне значимости 5% и проверить нулевую гипотезу на уровне доверия 95% – это одно и то же.

Логика проверки статистических гипотез

Предположим, что мы проводим исследование, посвященное степени поддержки разных кандидатов в губернаторы в регионе. У нас есть гипотеза, которая утверждает, что доля сторонников кандидата A в регионе равна 0.7. Мы опросили 1000 человек и выяснили, что доля респондентов, которые поддерживают кандидата A, равна 0.5. Можно ли сразу по таким результатам опроса сделать однозначный вывод, что доля сторонников кандидата в регионе не равна 0.7 (ведь $0.5 \neq 0.7$)? Нельзя.

Во-первых, мы уже знаем, что оценки параметра (в данном случае доли), полученные по одной выборке, отличаются от истинного значения параметра генеральной совокупности. Поэтому из того факта, что доля сторонников кандидата A в выборке равна 0.5, не следует, что доля его сторонников по всему региону обязательно равна 0.5.

Во-вторых, нам неизвестно, какая разница между выборочной долей и долей, заявленной в гипотезе, считается «маленькой», то есть достаточной для того, чтобы не отвергнуть нулевую гипотезу. В нашем примере доля сторонников кандидата A в выборке равна 0.5, мы можем считать, что 0.5 сильно отличается от 0.7, поэтому нам следует отвергнуть нулевую гипотезу. А что было бы, если бы выборочная доля была бы 0.6? Или 0.65? Сделали бы мы тогда вывод, что доля сторонников кандидата A в регионе не равна 0.7? Непонятно, потому что неизвестно, что считать сильным отличием, а что просто списывать на неточность оценок, получаемых по выборке.

Для того, чтобы понять, являются ли различия между значением в гипотезе и полученным по выборке, действительно существенными или эти различия – просто следствие того, что оценки по выборке мы получаем с некоторой погрешностью, требуется формальная проверка гипотез. Для разных видов вопросов существуют свои статистические критерии, позволяющие проверять соответствующие им нулевые гипотезы.

Статистический критерий – правило, которое позволяет делать вывод о том, стоит ли на основе имеющихся данных отвергать нулевую гипотезу или нет. Обычно для критерия определяется соответствующая ему **статистика** – мера отличия оценок параметров распределения генеральной совокупности, посчитанных по выборке, от тех значений, которые ожидаются в случае, если нулевая гипотеза верна. Статистика критерия имеет свое распределение. Для того чтобы понять, действительно ли разница между значением параметра в гипотезе и значением оценки, полученной по выборке, является существенной, вычисляют значение, которое называется **p-value**. Затем на основе полученного p-value делают вывод о том, стоит ли отвергать нулевую гипотезу или нет.

Концепция p-value

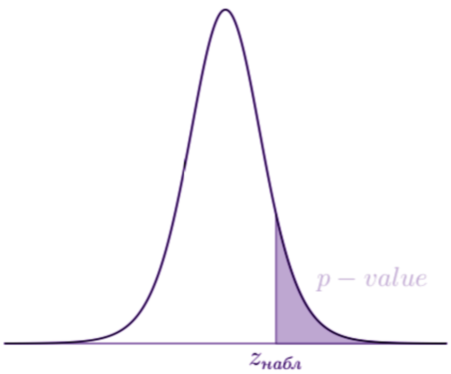
P-value – это вероятность получить значение статистики критерия равное наблюдаемому или более нетипичное по сравнению с наблюдаемым при условии, что нулевая гипотеза верна. Более неформально, p-value – это «жизнеспособность» нулевой гипотезы, которую мы оцениваем по имеющимся данным. Рассмотрим пример. У нас есть знакомый, который вдруг почувствовал себя нехорошо. Наша нулевая гипотеза заключается в том, что знакомый не болен, а просто переутомился перед сессией. Мы собираем различные данные, например, температуру тела, давление, другие симптомы, которые можно измерить количественно. По результатам анализа этих данных мы делаем вывод о том, что знакомый просто переутомился. Насколько мы оказались правы? Вероятность того, что значения показателей о здоровье, которые мы получили, действительно объясняются тем, что наш знакомый не болен, и есть p-value.

Что нам дают эти сведения? Во-первых, понимание, что p-value – это вероятность, причем не простая, а условная. Поэтому значения p-value всегда будут принадлежать интервалу $[0, 1]$. Во-вторых, так как p-value – это вероятность того, что нулевая гипотеза жизнеспособна, чем выше p-value, тем лучше, если мы хотим, чтобы нулевая гипотеза не была отвергнута.

Расчет p-value зависит от того, какого типа альтернативная гипотеза. Рассмотрим все случаи на примере z-статистики, которая имеет стандартное нормальное распределение.

- **Правосторонняя альтернативная гипотеза**
  Так как критическая область (область нетипичных значений статистики при условии, что нулевая гипотеза верна) находится справа, нас интересует площадь «хвоста» справа от наблюдаемого значения:

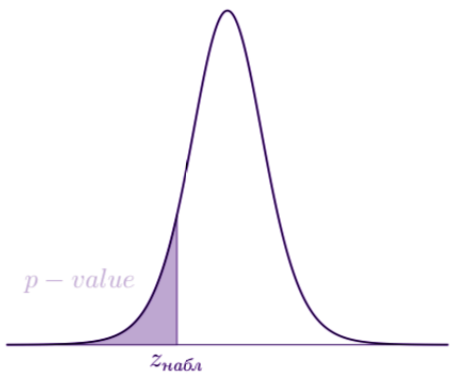
$$\text{p-value} = P(z \geq z_{\text{набл}}) = 1 - P(z < z_{\text{набл}})$$



- **Левосторонняя альтернативная гипотеза**
  Так как критическая область находится слева, нас интересует площадь «хвоста» слева от наблюдаемого значения:

$$\text{p-value} = P(z \leq z_{\text{набл}})$$

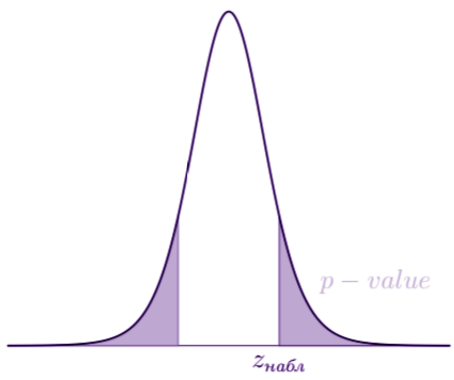
  Почему это так? Левостороннюю альтернативу мы выбираем в случае, когда есть основания считать, что истинное значение параметра будет меньше значения, указанного в нулевой гипотезе. Когда такие основания появляются? Когда выборочная оценка параметра (например, доля, посчитанная по выборке) меньше значения, зафикисированного в нулевой гипотезе. Все ещё более нетипичные значения статистики находятся в левом «хвосте»:



- **Двусторонняя альтернативная гипотеза**

$$\text{p-value} = P(|z| \geq z_{набл})$$

Почему это так? В случае двусторонней альтернативной гипотезы нас интересуют оба «хвоста» распределения, как правый, так и левый:



Как по p-value определить, есть ли основания отвергнуть нулевую гипотезу? Тут важно сначала зафиксировать уровень значимости α, а потом уже делать выводы. Уровень значимости α – это вероятность отвергнуть нулевую гипотезу при условии, что она верна. P-value – это минимальный уровень значимости, на котором нулевая гипотеза может быть отвергнута. Соответственно, если p-value меньше нашего фиксированного уровня значимости, на котором мы проверяем гипотезу, то нулевую гипотезу следует отвергнуть, если более – то отвергать нулевую гипотезу оснований нет. Если вдруг получилось, что p-value совпало с уровнем значимости (а это бывает довольно редко), то в таких случаях поступают на усмотрение исследователя.

Итак, получается:

- p-value $< \alpha \Rightarrow H_0$ отвергаем на уровне значимости α, на имеющихся данных
- p-value $> \alpha \Rightarrow H_0$ не отвергаем на уровне значимости α, на имеющихся данных

Важно: в выводе относительно отвержения / не-отвержения нулевой гипотезы необходимо указывать уровень значимости, так как от этого зависит результат. Так, например, в случае, если p-value равно 0.02, у нас есть основания отвергнуть нулевую гипотезу на уровне значимости 5% ($0.02 < 0.05$), и нет оснований отвергнуть ее на уровне значимости 1% ($0.02 > 0.01$).

Алгоритм проверки статистических гипотез

1. Сформулировать нулевую гипотезу (H_0).
2. Сформулировать альтернативную гипотезу (H_1).

   NB: Важно указывать, так как от типа альтернативной гипотезы (двусторонняя или односторонняя) зависит значение p-value.

3. Выбрать критерий, необходимый для проверки нулевой гипотезы.
4. Определить наблюдаемое значение статистики.
5. Посчитать p-value. Сравнить p-value с зафиксированным уровнем значимости.
6. Сделать статистический и содержательный вывод.

   Пример статистического вывода: на имеющихся данных, на уровне значимости 5% (уровне доверия 95%) есть основания/нет оснований отвергнуть нулевую гипотезу в пользу альтернативы.

   Пример содержательного вывода: среднее значение роста женщин не равно 168 см.

   NB: Важно всегда указывать уровень значимости (уровень доверия), на котором проверяется гипотеза, так как без этого уточнения выводы о нулевой гипотезе не имеют большого смысла: на одном уровне значимости гипотеза может быть отвергнута, а при выборе другого уровня значимости – нет. Желательно также прописывать, что выводы делаются на имеющихся данных, так как мы можем отвечать только за те результаты, которые получили по той выборке / выборкам, которые у нас есть, а не за «истинность» выводов вообще.

Важно! По результатам проверки статистической гипотезы мы никогда не делаем вывод о том, что нулевая гипотеза верна / должна быть принята. Вопрос об истинности нулевой гипотезы – содержательный вопрос, и если он и проверяется статистически, то с помощью более продвинутых методов и в рамках специально продуманного дизайна исследования. Всё, что мы можем решить по итогам проверки: отвергнуть нулевую гипотезу или нет. Как из того, что события не независимы, автоматически не следует их зависимость, так и из того, что нулевая гипотеза не отвергается, не следует, что она принимается.